Check for updates

Optimizing the OCTA layer fusion option for deep learning classification of diabetic retinopathy

Behrouz Ebrahimi,[1] **David Le,**[1] **Mansour Abtahi,**[1] **Albert K. Dadzie,**[1] **Jennifer I. Lim,**[2] **R. V. Paul Chan,**[2] **and Xincheng Yao**[1,2,*]

[1]*Department of Biomedical Engineering, University of Illinois at Chicago, Chicago, IL 60607, USA*
[2]*Department of Ophthalmology and Visual Sciences, University of Illinois at Chicago, Chicago, IL 60612, USA*
*xcy@uic.edu

Abstract: The purpose of this study is to evaluate layer fusion options for deep learning classification of optical coherence tomography (OCT) angiography (OCTA) images. A convolutional neural network (CNN) end-to-end classifier was utilized to classify OCTA images from healthy control subjects and diabetic patients with no retinopathy (NoDR) and non-proliferative diabetic retinopathy (NPDR). For each eye, three en-face OCTA images were acquired from the superficial capillary plexus (SCP), deep capillary plexus (DCP), and choriocapillaris (CC) layers. The performances of the CNN classifier with individual layer inputs and multi-layer fusion architectures, including early-fusion, intermediate-fusion, and late-fusion, were quantitatively compared. For individual layer inputs, the superficial OCTA was observed to have the best performance, with 87.25% accuracy, 78.26% sensitivity, and 90.10% specificity, to differentiate control, NoDR, and NPDR. For multi-layer fusion options, the best option is the intermediate-fusion architecture, which achieved 92.65% accuracy, 87.01% sensitivity, and 94.37% specificity. To interpret the deep learning performance, the Gradient-weighted Class Activation Mapping (Grad-CAM) was utilized to identify spatial characteristics for OCTA classification. Comparative analysis indicates that the layer data fusion options can affect the performance of deep learning classification, and the intermediate-fusion approach is optimal for OCTA classification of DR.

1. Introduction

Diabetic retinopathy (DR) is a leading cause of vision loss in developed countries [1]. If it could be diagnosed promptly, more than 95% of DR-related vision loss is preventable [2], at least the progress can be slowed through appropriate treatment [3–6]. Regular DR screening is recommended for all patients with diabetes [7]. DR can be characterized by various retinal vascular abnormalities, such as microaneurysms, retinal edema, hard exudates, retinal hemorrhages, venous beading, and intraretinal microvascular anomalies [8]. In addition to retinal vascular markers, choroidal abnormalities have been also reported in DR [9], and the choroidal blood flow deficit may represent an early pathologic alteration in DR [10].

Given the importance of early DR detection, it is crucial to detect subtle retinal and choroidal abnormalities. Traditional fundus photography has limitations in terms of sensitivity to reveal subtle abnormalities [11–14]. Fluorescein angiography (FA) can improve imaging sensitivity of retinal vascular distortions in DR [15]. However, it requires dye injections which can cause adverse side effects. In contrast, optical coherence tomography (OCT) angiography (OCTA) is a non-invasive imaging method that provides volumetric data of the retinal and choroidal layers [16,17], and has been demonstrated to be more sensitive than FA in detecting subtle abnormalities [18].

Quantitative OCTA features have been demonstrated for objective detection and classification of DR [19–24]. Machine learning and deep learning methods have been employed to classify DR automatically [25–28]. In principle, deep learning models can extract complex patterns and subtle features from unimodal or multimodal data. Some studies have combined multiple modalities to achieve better performance [29,30], while others have utilized fusion strategies within the same modality [31]. Because different OCTA layers may reflect different aspects of the ocular condition, it is important to test and optimize layer fusion options in deep learning detection and classification of eye diseases.

The potential data fusion strategies can be divided into three categories: early, intermediate, and late-fusion [32,33]. The early-fusion combines information at the raw data level before any processing takes place. Ryu et al. have utilized this approach by concatenating the superficial capillary plexus (SCP), deep capillary plexus (DCP), and full thickness retina layer of OCTA [31]. The intermediate-fusion combines information from each layer after feature extraction before they are fed into the final classification layer. The late-fusion combines layer information at the decision stage, after feature extraction and classification with individual layers. The late-fusion techniques have been inspired by ensemble learning approaches in a variety of ways [34].

Given the various data fusion strategies, this study aims to assess the effects of early, intermediate, and late-fusion options on OCTA classification of DR. In addition to retinal layers, additional choriocapillaris (CC) OCTA may help improving DR classification performance, due to DR caused choroidal abnormalities [9,10]. Thus, this study evaluated multi-layer fusion strategies in OCTA using en-face images from SCP, DCP, and CC for deep learning classification of DR. To assist with the interpretation of the classification performance, class activation maps were used to identify the most significant regions in each OCTA layer.

2. Methods

A convolutional neural network (CNN) classifier was utilized in this study. Early-fusion, intermediate-fusion, and late-fusion strategies were included to evaluate the effect of multi-layer approaches on the performance of DR classification. The following section 2.1 describes data acquisition. Section 2.2 delves into the CNN architecture and implementation. Section 2.3 discusses the classification using individual OCTA layers. Section 2.4 focuses on the combination of multiple OCTA layers via various fusion structures. Section 2.5 defines the evaluation metrics of deep learning OCTA classification of DR. Finally, section 2.6 describes class activation maps to visualize the important regions for DR classification.

2.1. Data acquisition

In this study, OCTA volumetric data from 136 subjects were acquired, including 46 healthy control eyes, 26 eyes of diabetic patients with no DR (NoDR), and 64 eyes of patients with non-proliferative DR (NPDR). These datasets were obtained from University of Illinois Chicago (UIC) eye clinic and were acquired using ANGIOVUE spectral domain (SD) OCTA systems (Optovue, Fremont, CA). These OCTA images were centered on the fovea and covered a 6×6 mm^2 area. As illustrated in Fig. 1, for each eye, three en-face images from the SCP, DCP, and CC layers were generated. This study was conducted in accordance with the ethical standards outlined in the Declaration of Helsinki and was approved by the institutional review board of UIC.

2.2. CNN classifier and implementation details

The image minimal and maximal intensities were normalized before being fed to the model. The base architecture chosen for this study was EfficientNetV2L [35]. As shown in Fig. 2, the CNN-based end-to-end classifier for DR classification can be divided into two parts: the first part extracts features from the OCTA images, and the second part uses these features to
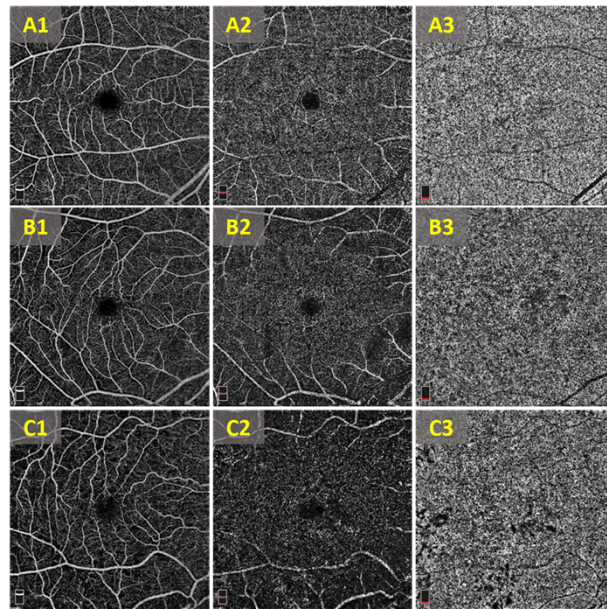
Fig. 1. (A) Representative OCTA images of SCP (A1), DCP (A2), and CC (A3) of a healthy control subject. (B) Representative OCTA images of SCP (B1), DCP (B2), and CC (B3) of a NoDR subject. (C) Representative OCTA images of SCP (C1), DCP (C2), and CC (C3) of an NPDR subject.

classify the images into the appropriate groups. The transfer learning technique was employed to compensate for the limited dataset size of available OCTA images [36]. Transfer learning is a training approach that uses some weights from a pretrained CNN to retrain specific layers of the network [37]. In order to avoid overfitting and enhance the generalizability, data augmentation operations, including random rotation, brightness change, horizontal and vertical flip, zooming, and scaling, were applied. The training was done for 200 epochs, with a learning rate of 0.00001, optimizer of Adam, loss function of categorical cross entropy, batch size of 32, and a callback function of early stopping. Because of the limited dataset size, the five-fold cross validation was performed. The network was trained using 80 percent of the images in each fold, and the remaining 20 percent of the images were utilized as validation.
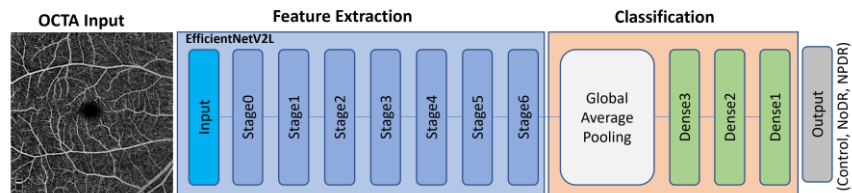

Fig. 2. CNN-based end-to-end classifier for DR classification

For all experiments except the late-fusion, the pretrained weights from the ImageNet dataset were transferred to the EfficientNetV2L base model [38]. In the late-fusion experiment, the pretrained weights from the individual OCTA layer models were utilized. The model was implemented using Python v3.8 software with the Keras 2.9.0 and TensorFlow 2.9.1 open-source

platform backend. Training was performed on a Windows 10 computer with an NVIDIA RTX 6000 Ti graphics processing unit.

2.3. CNN classifier for individual OCTA layer architectures

The ability of the CNN classifier to classify OCTA images into different groups was first assessed by training the model separately on each of the three OCTA layers. This allowed us to determine the most informative layer for DR classification and to compare the performance of the model on different layer fusion options. Figure 3 illustrates individual layer inputs (Fig. 3(A)) and multi-layer fusion (Fig. 3(B)) options. A five-fold cross-validation approach was used in the training of the model, where the data was divided into five equal sets and the training and evaluation of the model was carried out on each set. This allowed us to evaluate the performance of the model on a diverse set of images and to estimate the generalizability of the model to new data. The individual OCTA layer architecture is defined as SCP-only, DCP-only, or CC-only based on the input layer used. After training the model on individual layers, the combination of the layers in different approach to check the classification performance was examined. Three different fusion architectures: early-fusion, intermediate-fusion, and late-fusion were considered in this study.
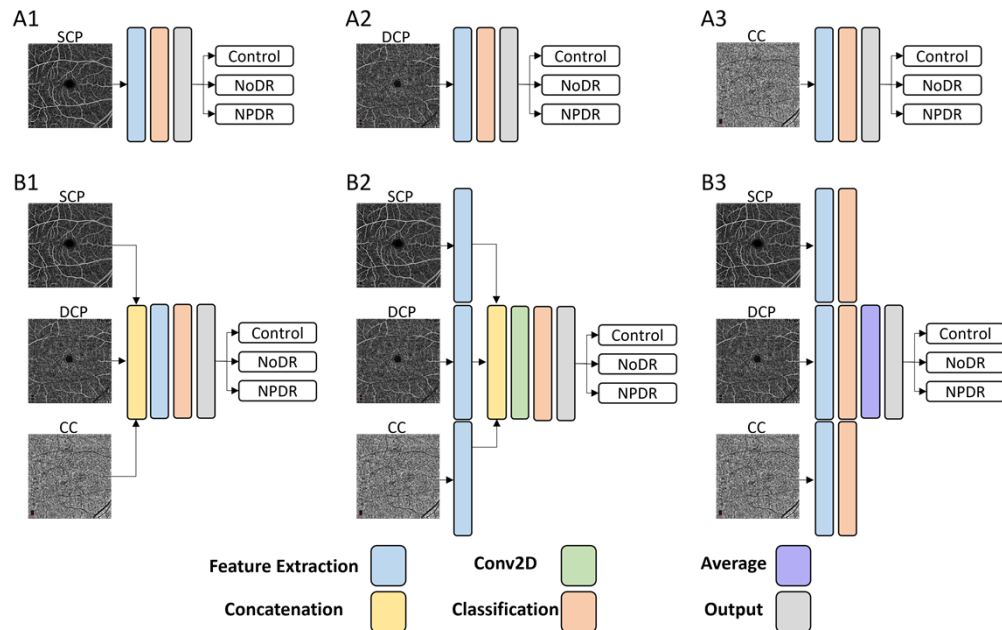
Fig. 3. (A) DR classification with SCP-only (A1), DCP-only (A2), and CC-only (A3) OCTA. (B) DR classification with early-fusion (B1), intermediate-fusion (B2), late-fusion (B3) architectures.

2.4. Early-fusion, intermediate-fusion, and late-fusion

2.4.1. Early-fusion

Early-fusion is a data fusion strategy that combines the raw data from different sources at the input level of the model [32]. In this study, early-fusion involves concatenating the raw data from the SCP, DCP, and CC layers of OCTA and presenting them to the model as three separate input channels (Fig. 3(B1)). By combining the raw data from different layers at the input level, the

model can learn to integrate and exploit the complementary information from the different layers to improve the classification performance.

2.4.2. Intermediate-fusion

Intermediate-fusion is a data fusion strategy that combines the processed data from different sources after some initial processing has been applied to the raw data [32]. In this study, intermediate-fusion combines features derived from the SCP, DCP, and CC layers for following processing and classification (Fig. 3(B2)). Each layer OCTA is first processed separately through a feature extraction module. The outputs of the feature extraction modules are then concatenated and fed into a convolutional layer, which further processes the data and extracts higher-level features. The output of the convolutional layer is then passed to the classification module to produce the final prediction.

2.4.3. Late-fusion

Late-fusion is a data fusion strategy to combine the different sources after all processing has been completed [32]. In this study, late-fusion involves applying this approach to the data from the SCP, DCP, and CC layers of OCTA. As depicted in Fig. 3(B3), this involves extracting features and performing classification separately for each layer, using the fully processed data from each input. The outputs of the classification modules for three layers are then combined using a global averaging layer. The final prediction is produced based on the combined outputs of all three layers.

Initially, the pretrained weights from the separate individual OCTA layer models (Fig. 3(A)), were employed, instead of using pretrained weights of ImageNet data set to the EfficientNetV2L base model [38]. This strategy was employed since the model was unable to converge. Due to the size of the model in late-fusion and the number of parameters to be trained, the model cannot be properly trained with a small dataset unless overfitting happens. To avoid this problem, the best starting point possible was required to initialize the parameters with the pretrained weights of each individual model.

2.5. Evaluation metrics of deep learning performance

Several metrics were employed in this study to evaluate the performance of the deep learning models and to quantify their accuracy and effectiveness. One common metric for evaluating the performance of classification models is the receiver operating characteristic (ROC) curve, which plots the true positive rate (sensitivity) against the false positive rate (1-specificity) at various classification thresholds. The area under the curve (AUC) of the ROC curve is another useful metric, as it measures the entire area underneath the ROC curve. A model with an AUC of 1 is a perfect classifier, while a model with an AUC of 0.5 is no better than random chance.

In addition to ROC and AUC, other performance metrics, including accuracy (ACC), sensitivity (SE), and specificity (SP), were also calculated. Sensitivity is a measure of the proportion of true positives that are correctly identified. Specificity is a measure of the proportion of true negatives that are correctly identified. Accuracy is a measure of the overall performance of the model. These metrics are defined as:

$$Sensitivity = \frac{TP}{TP+FN} \tag{1}$$

$$Specificity = \frac{TN}{TN+FP} \tag{2}$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{3}$$

where TP, TN, FP, and FN represent the number of true positives, true negatives, false positives, and false negatives, respectively.

2.6. Class activation map

The Gradient-weighted Class Activation Mapping (Grad-CAM) [39] was utilized to identify the OCTA regions that were most important for the classification decision. Specifically, the input image was fed into the pre-trained CNN and the gradient information flowing into the final convolutional layer was used to generate a class activation map. The class activation map was then overlaid onto the input image to create a heatmap visualization, which highlights the image regions that were most influential for the classification decision.

3. Results

The cross-validation performances for both individual layer inputs and multi-layer fusion architectures are summarized in Fig. 4 and Table 1. For individual layer inputs, the SCP-only provided the best performance, with the highest accuracy (87.249%), sensitivity (78.264%), and specificity (90.104%). For multi-layer fusion architectures, the intermediate-fusion architecture showed the best performance. The early-fusion showed a slightly lower performance compared to the SCP-only architecture. The late-fusion architecture had slightly worse results compared to intermediate-fusion, but still outperformed all other options.
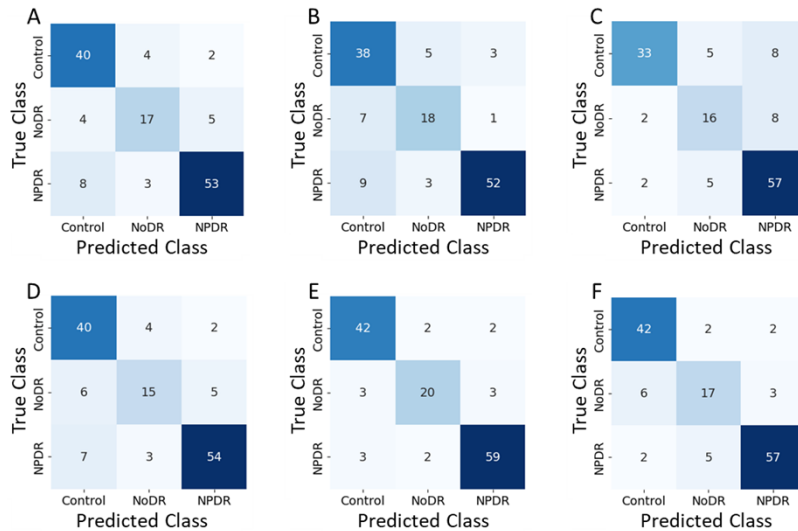
Fig. 4. Confusion matrices of SCP-only (A), DCP-only (B), CC-only (C), early-fusion (D), intermediate-fusion (E), and late-fusion (F) architectures.

Figure 5 illustrates ROC curves of different groups in different architectures. The overall AUC for SCP-only, DCP-only, CC-only, early-fusion, intermediate-fusion, and late-fusion are 0.894, 0.857, 0.851, 0.875, 0.916, and 0.915, respectively. Among all architectures, the intermediate-fusion architecture has the highest AUC value, which is consistent with the findings in Table 1. For individual OCTA layer inputs, the SCP-only and DCP-only obtained higher AUC values of the NPDR group than the control and NoDR groups, while the CC-only achieved a higher AUC value of the control group than the NoDR and NPDR groups. For multi-layer fusion architectures, the intermediate-fusion and the early-fusion obtained higher AUC values of the NPDR group than the control and NoDR groups, whereas the late-fusion achieved a higher AUC value of the control group than the NoDR and NPDR groups. For all individual OCTA layer and multi-layer fusion architectures, the NoDR group attained a lower AUC value than the control and NPDR groups.

**Table 1. Comparative performance illustration of OCTA layer fusion options**

Architectures	Class	Metrics			
		Acc (%)	Se (%)	Sp (%)	AUC
SCP-only	Control	86.772	87.111	86.588	0.8971
	NoDR	88.228	65.333	93.546	0.8752
	NPDR	86.746	82.348	90.179	0.9121
	Average	87.249	78.264	90.104	0.8948
DCP-only	Control	82.381	83.111	82.384	0.8449
	NoDR	88.228	68.667	92.554	0.8444
	NPDR	88.228	80.919	94.643	0.8839
	Average	86.279	77.566	89.860	0.8577
CC-only	Control	87.566	72.222	95.607	0.9179
	NoDR	85.291	61.333	90.775	0.7850
	NPDR	83.069	88.951	77.679	0.8522
	Average	85.309	74.169	88.020	0.8517
Early-fusion	Control	86.085	86.889	85.587	0.9019
	NoDR	86.772	57.333	93.589	0.7958
	NPDR	87.460	83.716	90.179	0.9282
	Average	86.772	75.980	89.785	0.8753
Intermediate-fusion	Control	92.646	91.556	93.326	0.9220
	NoDR	92.672	77.333	96.399	0.8776
	NPDR	92.619	92.138	93.393	0.9494
	Average	**92.646**	**87.009**	**94.373**	**0.9163**
Late-fusion	Control	91.217	91.333	91.163	0.9442
	NoDR	88.228	65.333	93.676	0.8598
	NPDR	91.138	88.951	93.393	0.9431
	Average	90.194	81.873	92.744	0.9157

Figure 6 shows representative images to demonstrate the class activation maps of an NPDR OCTA. For individual layer inputs, the SCP-only (Fig. 6(A1)) and the DCP-only (Fig. 6(A2)) models were observed to involve the foveal avascular zone (FAZ) and surrounding capillary dropouts for the classification. The CC-only model (Fig. 6(A3)) also revealed dropout areas with reduced blood flow and vascular density.

By comparing the intermediate-fusion (Fig. 6(C)) and late-fusion (Fig. 6(D)), it was observed that both fusion strategies preserve the features learned from the separate OCTA layers, and the layer information is combined to bolster the classification performance. Whereas the early-fusion (Fig. 6(B)) model cannot completely discern the specific layer information for DR classification, thereby having less consistent results. Specifically, the DCP layer (Fig. 6(B2)) in the early-fusion model failed to focus on the FAZ and surrounding areas, unlike the intermediate-fusion and late-fusion models.
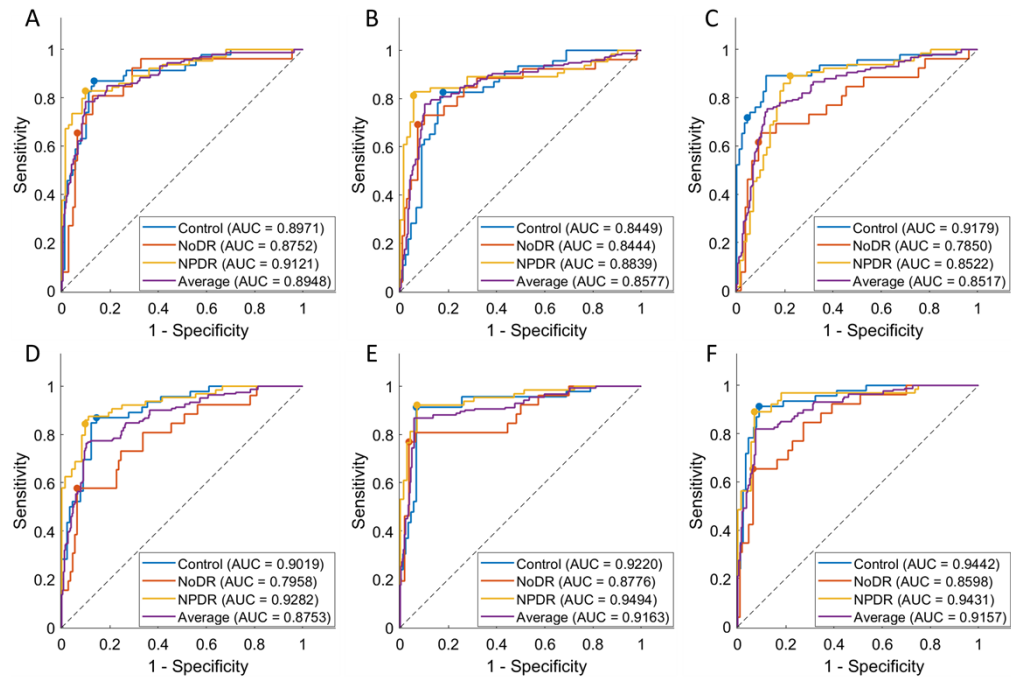
Fig. 5. ROC curves of SCP-only (A), DCP-only (B), CC-only (C), early-fusion (D), intermediate-fusion (E), and late-fusion (F) architectures.
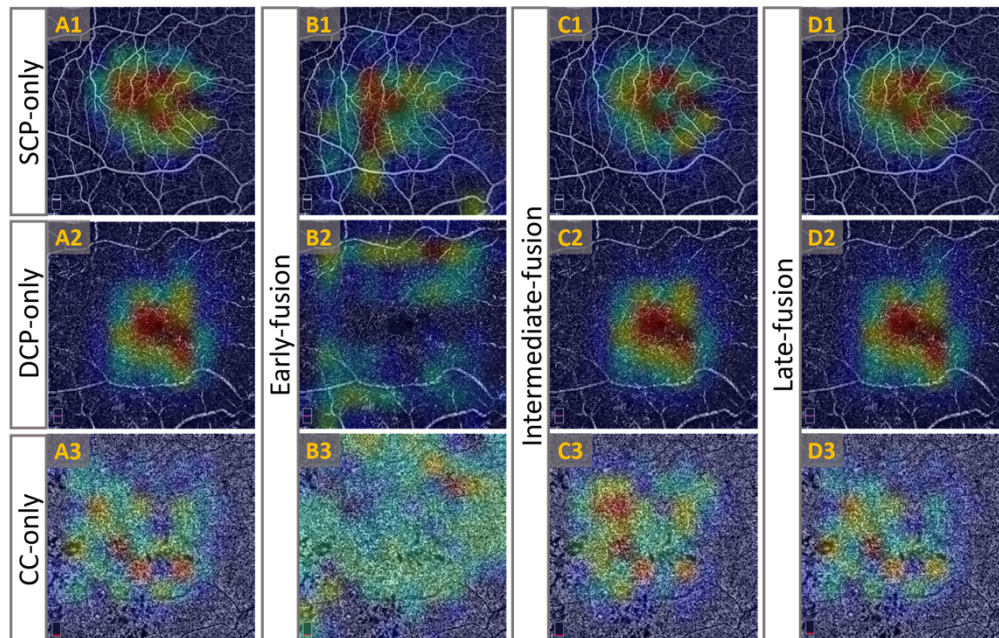
Fig. 6. Representative Grad-CAM results for an NPDR patient to highlight the regions useful for deep learning classification in individual layer inputs (A), early-fusion (B), intermediate-fusion (C), and late-fusion (D).

4.　Discussion

In this study, the performances of individual OCTA layer inputs and multi-layer fusion archi-tectures, including early-fusion, intermediate-fusion, and late-fusion were evaluated for deep learning classification of DR. For individual OCTA layer inputs, the SCP-only architecture achieved the best performance over the DCP-only and CC-only architectures. For the multi-layer architectures, the intermediate-fusion architecture had the best performance over the late-fusion and early-fusion architectures.

We speculated that combined information of OCTA layers would help robust performance of deep learning classification of DR. However, this study indicates that the early-fusion approach did not improve the performance compared to the SCP-only architecture. There are several possibilities to explain the observation. First, the early-fusion approach may not be able to identify correlations between the OCTA layers. In fact, by combining the raw data directly, the deep learning model may not be able to discriminate different layers and then their pathological correlations [32]. In other words, the pathological features of these layers may only become apparent when the model is able to consider them at a more abstract level, i.e., deeper in the CNN. Therefore, with the early-fusion approach, it may not be possible for the CNN to perform optimally. Ryu et al. utilized a full thickness retina layer as a third input and observed slightly improved accuracy compared to a single layer input [31]. At the raw data level, the full thickness retina layer demonstrates a stronger correlation with SCP and DCP than CC. Second, the intensity values and spatial patterns of input images may be different, which can make it difficult for the model to optimize its weights during the training process [40]. The CC layer OCTA is brighter than the other two layers. Therefore, if the brightness values are very different, it may be difficult for the model to optimize the weights of the first convolutional layer, which is responsible for merging these layers. For this study, we have normalized the image brightness, according to the range of minimal and maximal intensities, to minimize the effect of brightness variance. However, the signal to noise ratio and spatial pattern distribution of individual layers might be still different, and then can negatively impact the early-fusion performance.

Unlike the early-fusion approach, the intermediate-fusion approach allows the model to first extract features from each OCTA layer separately. This separation allows the model to consider the unique characteristics of each layer and identify patterns and relationships that might not be visible when all layers are combined at an early stage. The extracted features on each layer are then concatenated together and passed to the convolutional and dense layers of the model. This architecture enables the fusion at an appropriate level of abstraction and understand how the different layers are related to each other. As a result, this whole process enables the model to consider the higher representation of each layer and find correlations between them (Fig. 3(B2)).

In contrast, the late-fusion approach fuses the OCTA layers at the last stage, which may not allow the model to fully take advantage of the correlations among the layers. As a result, the performance of the model is slightly worse than intermediate-fusion, but still better than early-fusion and individual layer inputs. Comparing the late-fusion and early-fusion, Heisler et al. demonstrated that combining superficial and deep plexus enface images of OCTA and deep plexus enface images of OCT at the end stage provides higher performance than combining at the early stage using standard CNN [41].

Deep learning is widely known for its exceptional performance in various classification tasks, but the lack of interpretability due to the automated feature learning and extraction process has been a significant challenge. To address this issue, various methods have been proposed, among which the Grad-CAM approach has gained popularity [39]. In Fig. 6 we demonstrate the region of interest for an NPDR patient in different models. Based on these qualitative results, the intermediate-fusion (Fig. 6(C)) and late-fusion (Fig. 6(D)) are able to preserve the features learned from individual OCTA layer inputs (Fig. 6(A1), (A2), and (A3)), resulting in high and comparable classification performance. In contrast, the early-fusion model failed to focus on important

regions, leading to lower classification performance. Specifically, the early-fusion model missed the FAZ area in the DCP layer (Fig. 6(B2)), whereas the intermediate-fusion and late-fusion models were able to capture this important feature (Fig. 6(C2), and (D2)). These findings suggest that the intermediate-fusion and late-fusion models can complement the information learned from separate OCTA layers, resulting in improved performance compared to the early-fusion model.

5. Conclusion

Comparative analysis indicates that the deep learning performance can be significantly affected by the layer fusion options. For individual OCTA layer inputs, the SCP-only model showed the best performance for DR classification. For multi-layer fusion options, the intermediate-fusion achieved the best performance. The presented Grad-CAMs showed that the intermediate-fusion and late-fusion can preserve the features learned from individual OCTA layers to bolster the classification performance. On the contrary, the early-fusion model cannot effectively identify layer correlations for robust classification.

Disclosures. No competing interest exists for any author.

Data availability. Data may be obtained from the authors upon reasonable request.

References

1. M. M. Nentwich and M. W. Ulbig, "Diabetic retinopathy-ocular complications of diabetes mellitus," World journal of diabetes **6**(3), 489 (2015).
2. N. M. Glasson, L. J. Crossland, and S. L. Larkins, "An innovative Australian outreach model of diabetic retinopathy screening in remote communities," J. Diabetes Res. **2016**, 1–10 (2016).
3. M. J. Elman, L. P. Aiello, R. W. Beck, N. M. Bressler, S. B. Bressler, A. R. Edwards, F. L. Ferris III, S. M. Friedman, A. R. Glassman, and K. M. Miller, "Randomized trial evaluating ranibizumab plus prompt or deferred laser or triamcinolone plus prompt laser for diabetic macular edema," Ophthalmology **117**(6), 1064–1077.e35 (2010).
4. P. Massin, F. Bandello, J. G. Garweg, L. L. Hansen, S. P. Harding, M. Larsen, P. Mitchell, D. Sharp, U. Wolf-Schnurrbusch, and M. Gekkieva, "Safety and Efficacy of Ranibizumab in Diabetic Macular Edema (RESOLVE Study) A 12-month, randomized, controlled, double-masked, multicenter phase II study," Diabetes care **33**(11), 2399–2405 (2010).
5. M. Michaelides, A. Kaines, R. D. Hamilton, S. Fraser-Bell, R. Rajendram, F. Quhill, C. J. Boos, W. Xing, C. Egan, and T. Peto, "A prospective randomized trial of intravitreal bevacizumab or laser therapy in the management of diabetic macular edema (BOLT study): 12-month data: report 2," Ophthalmology **117**(6), 1078–1086.e2 (2010).
6. P. Mitchell, F. Bandello, U. Schmidt-Erfurth, G. E. Lang, P. Massin, R. O. Schlingemann, F. Sutter, C. Simader, G. Burian, and O. Gerstner, "The RESTORE study: ranibizumab monotherapy or combined with laser versus laser monotherapy for diabetic macular edema," Ophthalmology **118**(4), 615–625 (2011).
7. D. A. Antonetti, R. Klein, and T. W. Gardner, "Mechanisms of disease diabetic retinopathy," N. Engl. J. Med. **366**(13), 1227–1239 (2012).
8. J. Nayak, P. S. Bhat, U. Acharya, C. M. Lim, and M. Kagathi, "Automated identification of diabetic retinopathy stages using digital fundus images," J Med Syst **32**(2), 107–115 (2008).
9. I. Gendelman, A. Y. Alibhai, E. M. Moult, E. S. Levine, P. X. Braun, N. Mehta, Y. Zhao, A. Ishibazawa, O. A. Sorour, and C. R. Baumal, "Topographic analysis of macular choriocapillaris flow deficits in diabetic retinopathy using swept–source optical coherence tomography angiography," Int J Retin Vitr **6**(1), 6–8 (2020).
10. H. Wang and Y. Tao, "Choroidal structural changes correlate with severity of diabetic retinopathy in diabetes mellitus," BMC Ophthalmol. **19**(1), 186 (2019).
11. K. R. Mendis, C. Balaratnasingam, P. Yu, C. J. Barry, I. L. McAllister, S. J. Cringle, and D.-Y. Yu, "Correlation of histologic and clinical images to determine the diagnostic value of fluorescein angiography for studying retinal capillary detail," Invest. Ophthalmol. Visual Sci. **51**(11), 5864–5869 (2010).
12. S. Zahid, R. Dolz-Marco, K. B. Freund, C. Balaratnasingam, K. Dansingani, F. Gilani, N. Mehta, E. Young, M. R. Klifto, and B. Chae, "Fractal dimensional analysis of optical coherence tomography angiography in eyes with diabetic retinopathy," Invest. Ophthalmol. Visual Sci. **57**(11), 4940–4947 (2016).
13. B. I. Gramatikov, "Modern technologies for retinal scanning and imaging: an introduction for the biomedical engineer," BioMed Eng OnLine **13**(1), 52 (2014).
14. A. Rossi, M. Rahimi, D. Le, T. Son, M. J. Heiferman, R. P. Chan, and X. Yao, "Portable widefield fundus camera with high dynamic range imaging capability," Biomed. Opt. Express **14**(2), 906–917 (2023).

15. Ş. Ţălu, D. M. Călugăru, and C. A. Lupaşcu, "Characterisation of human non-proliferative diabetic retinopathy using the fractal analysis," International journal of ophthalmology **8**, 770 (2015).

16. K. Chalam and K. Sambhav, "Optical coherence tomography angiography in retinal diseases," J. Ophthalmic Vision Res. **11**(1), 84 (2016).

17. D. Le, T. Son, T. Kim, T. Adejumo, M. Abtahi, S. Ahmed, A. Rossi, B. Ebrahimi, A. Dadzie, and X. Yao, "SVC-Net: A spatially vascular connectivity network for deep learning construction of microcapillary angiography from single-scan-volumetric OCT," (2023).

18. S. Mo, B. Krawitz, E. Efstathiadis, L. Geyman, R. Weitz, T. Y. Chui, J. Carroll, A. Dubra, and R. B. Rosen, "Imaging foveal microvasculature: optical coherence tomography angiography versus adaptive optics scanning light ophthalmoscope fluorescein angiography," Invest. Ophthalmol. Visual Sci. **57**(9), OCT130 (2016).

19. D. Le, M. Alam, B. A. Miao, J. I. Lim, and X. Yao, "Fully automated geometric feature analysis in optical coherence tomography angiography for objective classification of diabetic retinopathy," Biomed. Opt. Express **10**(5), 2493–2503 (2019).

20. A. K. Dadzie, D. Le, M. Abtahi, B. Ebrahimi, T. Son, J. I. Lim, and X. Yao, "Normalized Blood Flow Index in Optical Coherence Tomography Angiography Provides a Sensitive Biomarker of Early Diabetic Retinopathy," arXiv, >arXiv:2212.14840 (2022).

21. Y.-T. Hsieh, M. N. Alam, D. Le, C.-C. Hsiao, C.-H. Yang, D. L. Chao, and X. Yao, "OCT angiography biomarkers for predicting visual outcomes after ranibizumab treatment for diabetic macular edema," Ophthalmology Retina **3**(10), 826–834 (2019).

22. M. Alam, Y. Zhang, J. I. Lim, R. V. Chan, M. Yang, and X. Yao, "Quantitative optical coherence tomography angiography features for objective classification and staging of diabetic retinopathy," Retina **40**(2), 322–332 (2020).

23. M. Alam, D. Le, J. I. Lim, and X. Yao, "Vascular complexity analysis in optical coherence tomography angiography of diabetic retinopathy," Retina **41**(3), 538–545 (2021).

24. M. Abtahi, D. Le, B. Ebrahimi, A. K. Dadzie, J. I. Lim, and X. Yao, "An open-source deep learning network AVA-Net for arterial-venous area segmentation in optical coherence tomography angiography," Commun. Med. **3**(1), 54 (2023).

25. P. Zang, L. Gao, T. T. Hormel, J. Wang, Q. You, T. S. Hwang, and Y. Jia, "DcardNet: Diabetic retinopathy classification at multiple levels based on structural and angiographic optical coherence tomography," IEEE Trans. Biomed. Eng. **68**(6), 1859–1870 (2021).

26. Z. Liu, C. Wang, X. Cai, H. Jiang, and J. Wang, "Discrimination of diabetic retinopathy from optical coherence tomography angiography images using machine learning methods," IEEE Access **9**, 51689–51694 (2021).

27. M. M. Abdelsalam and M. Zahran, "A novel approach of diabetic retinopathy early detection based on multifractal geometry analysis for OCTA macular images using support vector machine," IEEE Access **9**, 22844–22858 (2021).

28. J. Cano, W. D. O'neill, R. D. Penn, N. P. Blair, A. H. Kashani, H. Ameri, C. L. Kaloostian, and M. Shahidi, "Classification of advanced and early stages of diabetic retinopathy from non-diabetic subjects by an ordinary least squares modeling method applied to OCTA images," Biomed. Opt. Express **11**(8), 4666–4678 (2020).

29. Y. Li, M. El Habib Daho, P.-H. Conze, H. Al Hajj, S. Bonnin, H. Ren, N. Manivannan, S. Magazzeni, R. Tadayoni, and B. Cochener, "Multimodal information fusion for glaucoma and diabetic retinopathy classification," in *International Workshop on Ophthalmic Medical Image Analysis*(Springer2022), pp. 53–62.

30. Á. S. Hervella, J. Rouco, J. Novo, and M. Ortega, "Retinal microaneurysms detection using adversarial pre-training with unlabeled multimodal images," Information Fusion **79**, 146–161 (2022).

31. G. Ryu, K. Lee, D. Park, I. Kim, S. H. Park, and M. Sagong, "A deep learning algorithm for classifying diabetic retinopathy using optical coherence tomography angiography," Trans. Vis. Sci. Tech. **11**(2), 39 (2022).

32. S. R. Stahlschmidt, B. Ulfenborg, and J. Synnergren, "Multimodal deep learning for biomedical data fusion: a review," Briefings Bioinf. **23**(2), bbab569 (2022).

33. D. Ramachandram and G. W. Taylor, "Deep multimodal learning: A survey on recent advances and trends," IEEE Signal Process. Mag. **34**(6), 96–108 (2017).

34. L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms* (John Wiley & Sons, 2014).

35. M. Tan and Q. Le, "Efficientnetv2: Smaller models and faster training," in *International Conference on Machine learning* (PMLR2021), pp. 10096–10106.

36. D. Le, M. Alam, C. K. Yao, J. I. Lim, Y.-T. Hsieh, R. V. Chan, D. Toslak, and X. Yao, "Transfer learning for automated OCTA detection of diabetic retinopathy," Trans. Vis. Sci. Tech. **9**(2), 35 (2020).

37. D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, and F. Yan, "Identifying medical diagnoses and treatable diseases by image-based deep learning," Cell **172**(5), 1122–1131.e9 (2018).

38. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*(Ieee2009), pp. 248–255.

39. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*(2017), pp. 618–626.

40. M. Abtahi, D. Le, J. I. Lim, and X. Yao, "MF-AV-Net: an open-source deep learning network with multimodal fusion options for artery-vein segmentation in OCT angiography," Biomed. Opt. Express **13**(9), 4870–4888 (2022).

41. M. Heisler, S. Karst, J. Lo, Z. Mammo, T. Yu, S. Warner, D. Maberley, M. F. Beg, E. V. Navajas, and M. V. Sarunic, "Ensemble deep learning for diabetic retinopathy detection using optical coherence tomography angiography," Trans. Vis. Sci. Tech. **9**(2), 20 (2020).